The PLAUSIBLE Mutation of DNA

Consider first the analogy between a DNA molecule and a computer program. Transfer RNA "swaps in" the DNA "program", and at the ribosomes it is "EVAL'ed" (messenger RNA brings the required types of "freelist cells"). The "output" is a polypeptide chain (protein). The famous "genetic code" is the key with which triples of base pairs are converted into amino acids. That is the programming language's basic "Print" statement. Simple loop termination and other regulatory actions are brought about by the program -- the DNA: regulatory genes (which synthesize enzymes), insertion sequences, transposons, phage Mu, and other controlling elements. The analogy could be extended even further.

The DNA "program" develops (is improved and lengthened) by Evolution. That is, random changes occur in the sequence, manifest themeselves as mutated progeny, and are judged by Natural Selection. The DNA program for even such a complex organism as Man is assumed to have developed by such a random generate & test progression.

We in AI know only too well the weakness of doing automatic programming by random changes of (and random additions of new) program instructions. Certainly it CAN be done, but it is extremely slow. The AI answer is to add knowledge: add a collection of expert rules for programming in general and for the program's task domain in particular. Code synthesis and transformation is now done acording to these rules. While far from complete or foolproof, they are nevertheless far superior to blind changes in program instructions.

Idea #1: Can we extend the DNA==program analogy by somehow adding knowledge to the DNA, knowledge about which kinds of mutations are plausible, which kinds have been tried unsuccessfully, etc. That is, can we imagine what it might mean to turn DNA's random generator (random mutations in the next generation) into a plausible move generator? If there is a way to encode such knowledge, such heuristic guidance rules, then we might expect that an organism with that kind of compiled hindsight would evolve in much more regular, rapid a fashion. The "test" would still be natural selection, but instead of blind generation the DNA would be conducting (and recording) plausible experiments.

What would such heursitics "look like"; i.e., how might they be "implemented" in the DNA program? They could be written in the alphabet of bases, but their interpretation wouldn't be as codons for proteins. So someone (e.g., mRNA) would have to detect such heuristics and not copy them; or else at the ribosome they would have to be skipped over. At translation time, they would be NO-OPs. At times of reproduction, however, they would specify allowable (and prevent disallowed) changes to be made in the new copy. I.e., they would sanction certain complex copying "errors". The "left hand sides" of such heuristics could be almost completely specified by position (proximity to genes which they referred to in the rule), and the start of such a heuristic would have to be signalled by some special sequence of bases (much like parentheses in Lisp). Each heuristic would have some demarcated domain or scope.

Idea #2: Nature might already have become as good at programming as we have. DNA might have ALREADY evolved from random generate & test into an expert program (expert at mutating itself in plausible ways). The

recently-observed "introns" are non-coding regions of DNA which just might correspond to the above heuristics. Since they are hypothesized by us to be heuristics for dealing with DNA subsequences, and they themselves are also DNA subsequences, they (or at least SOME of them) might be able to modify, enlarge, improve themselves / each other.

What I conjecture is that Nature (= natural selection) began with primitive organisms and a random-mutation scheme for improving them. By this weak method (random generation, followed by stringent testing), the first primitive introns (heuristics) accidentally came into being. They immediately overshadowed the less efficient random-mutation mechanism, just as oxidation quickly dominated fermentation once it evolved.

Each heuristic proposes a plausible change (call it C) in the DNA. The progeny which incorporate C (call them PC) also get a new heuristic indicating that that kind of change has been made and is good. The progeny P which do not incorporate C also get a heuristic added, but this one says that a change of type C was tried and failed. If one group (P or PC) dominates the other, then that group's new heuristic will have proven to be correct. "False" heuristics die out with the organisms that contain them.

As the species evolves, so do the heuristics. One big lesson from AM was the NEED for new heuristics to evolve continuously. Otherwise, as animals got more and more sophisticated, they would begin to evolve more and more slowly (random mutations, or those guided by a fixed set of heuristics, would become less and less frequently beneficial to the complex organism). Until Eurisko was conceived, this would have been the end of the story. We would guess that new heuristics evolve randomly, and in the rare cases that they are improvements, they get perpetuated by the progeny which have them. Thanks to Eurisko, we see that since the heuristics are represented just like any other DNA, they can work on themselves as well: they can suggest plausible (and/or warn of classes of implausible) changes to make in both (i) the DNA which synthesizes proteins, and (ii) the DNA which serves as heuristics.

Phenomena accounted for by this hypothesis include: the biological function of introns [heuristics]; the rapid evolution of man in general and his brain in particular (much more rapid than one could expect from straight random mutation) [heuristic exploration instead of random trial and error]; the ABC result (mutation rate per gram of DNA is not constant, but rather is proportional to the lengths of the DNA molecules making up the sample) [mutations are mediated by the introns, whose relative number increases in proportion to DNA length (roughly)]; the Schimke result (relearning a mutation is much quicker than initial learning, and the intermediate state of the de-learned DNA is slightly larger than the original length) [the learning causes a new heuristic to form, and even after the mutation is forced to be un-learned, the heuristic which summarizes that experience remains]; the apparent increase in introns as one ascends the evolutionary ladder [more heuristics evolved]; the large morphological advances of some species (like Man) compared with others (like chimps and even more dramatically frogs), even though at the DNA sequence level they both advanced an equal number of base mutations [programs with more heuristics can get more done in N cpu cycles].

I called this a hypothesis, and shall now try to justify that claim. This has several aspects, which are treated in turn below.

Toward a Theory of what the DNA "Program" has Evolved Into

A reiteration of the central hypothesis:
DNA has evolved into an expert program, i.e., one with heuristics (the introns) for suggesting which (families of) mutations are (im)plausible. Since the introns are represented exactly the same as any other DNA, the introns can refer to (and operate on) themselves (in addition to referring to protein-encoding DNA). As species evolve viably, the body of heuristics is gradually altered (by updating and by the addition of new heuristics) to capture the additional history, to compile the new hindsight.

> What does this hypothesis "explain" that old ones don't?
> > First, this proposes a use for the introns.
> > > There must be SOME vital use, if we believe in the ubiquity and severity of natural selection.
> > > It fits data accumulated about introns (e.g., why the percentage of introns increases with the complexity of the organism).
> > Second, it explains how organisms can continue to evolve rapidly and effectively, even as their complexity grows to that of Man.
> > > It is a mechanism which may be sufficiently better than random mutation so as to lead to Man much quicker.
> > > It might explain, also, why man's brain evolved so rapidly
> > > > > 500 grams in 500,000 years (20k generations) is a big enlargement
> > Third, it could explain various nonuniformities in the rate of sequence evolution
> > > Though this is not as crucial as the previous two points Because (as Wilson, Carlson & White note): The speed at which an organism morphologically evolves seems totally unrelated to the rate at which his individual proteins (DNA base sequences) evolve. "This result raises doubts about the relevance of sequence evolution to the evolution of organisms".
> > > On the other hand, the REASON that some species evolve morphologically quickly can be attributed to their effective heuristics. Frogs, e.g., have poor heuristics and have not evolved much in eons. WC&W: "Since humans and chimps had a common ancestor, much more phenotypic change has occurred in the human lineage than in that of the chimpanzee... In spite of having evolved at an unusually high organismal rate, the human lineage does not appear to have undergone accelerated sequence evolution". So human heuristics are superior to chimps'; even though the evolutionary clock has ticked away the same number of sequence mutations, the humans have used their time better than chimps, and much better than frogs.

Anyway, here are some of the other "explainable" nonuniformities:

> > > Why some proteins evolve at rates 10 times as slow as others, yet the rate of evolution is almost constant for proteins within certain classes. As Wilson, Carlson, & White say (Biochem. Evolution, An.Rev. Biochem. 1977): "It has been hard to understand why the rate is steady within a given class. As explanations involving pos. natural selection did not seem satisfactory, some workers proposed a non-darwinian explanation... of the evolutionary clock..."
> > > > Our "explanation" is simply that the evolution is heuristically guided. Uniformity is demanded by randomness, not by intelligence.

> > > Why some parts of a protein (some amino acids, usually about 5%) are absolutely stable (NEVER appear to have undergone substitution even during long evolutionary time periods. (Cavalli p.741)
> > > > We posit that this is the recommendation of some heuristics.
> > > Why the mutation rate per gene is proportional to the total length of the DNA molecule, not a constant (ABC paper)
> > > > We propose that the mechanism for mutation is primarily under direct control of the introns. Thus, a random change in an intron subsequence is much more likely to have morphological consequences than the change of one base in an extron subsequence. Since the relative amount of introns is increasing with DNA length, so is the chance of hitting an intron, hence so is the rate of mutations per gram of DNA.

> What evidence led to THIS hypothesis, rather than some other?
> > The empirical necessity of doing automatic programming (and complex tasks as a whole) by HPP methods, not weak ones.
> > The painful way in which I was forced to build Eurisko's heuristics as concepts. I would not have suffered this had it not been necessary (i.e., selected for).
> > > In other words: a strong analogy to the progression of paradigms (at least, MY personal mental world views) in AI research (No-Heuristics --> GPS --> Dendral --> AM --> Eurisko)
> > Such appeals to analogy are not uncommon in molecular genetics
> > > Enzyme induction mechanisms were debated in terms of locks & keys, templates & forms, and other real-world images.
> > > Adaptors were conceived as analogues of electrical wire or pipe adaptors.
> > > The analogy of restriction enzyme action to text editing has been fruitful.
> > > Biologists would not have the HPP, let alone AM, let alone Eurisko, designs to draw upon for analogy, hence might take a long time to figure out what's going on (if DNA really HAS become an "expert program").
> > The simulation of what a discoverized MOLGEN might act like
> > > In particular, extending the analogy of DNA = = Programs
> > The idea that computer scientists might consciously, intelligently re-design a basis for life (or at least improve on the existing design)
> > > E.g., writing a program that was cleaner and more powerful than current DNA style
And then implement that program in wetware
> > > And the shock of realizing that Nature might already have become as good at programming as we have.

> What predictions can be made, assuming this hypothesis?
> > We want the most radical and unexpected ones, to test the hyp. We also want ones for which experiments can be readily executed.
> > One prediction is that the introns will increase slowly with time, within a species, as well as quickly as one crosses species boundaries.
> > > We should try to measure introns in fossils, if possible
> > > We should measure amounts of introns vs extrons in as many different species as possible, to see if the ratio increases monotonically with height on th evolutionary ladder.
> > > > Experiments to test this kind of thing are rapidly becoming readily performable, and will be performed.
> > > > As pointed out earlier, there is already weakly confirming evidence for this hypothesis:
> > > > No introns observed yet in prokaryotes
> > > > A single 14-base non-coding region is spliced out of

yeast. This is the most primitive intron.

> > > > In Drosophila, the 28s gene has several introns and is never transcribed.

> > > > In chick albumen, the ratio of introns/extrons is much higher.

> > We predict that there will be some kind of parenthesization to indicate the scope of the introns.

> > > One way this might appear is if the introns all began with a special short base sequence, or two, and perhaps multiple copies of that base sequence.

> > > Yesterday, Doug Brutlag told me that GAA and GGAA commonly occur at the front end of introns. These may be the [and (.

> > Another prediction is that introns might be usable across species boundaries. I.e., introns from humans might be very useful to mice.

> > > If we can crack the intron "code" (which may involve positional referents and straight history, as well as domain-independent heuristics) just a little, we can try to transfer some of the introns from an advanced organism into a primitive one. If we succeed, the subsequent generations of that organism should evolve MUCH faster than they otherwise would have, and probably in the direction of whatever the higher organism was.

> > > We expect that taking an intron located near an extron coding for protein P would be usable if place in the same proximity to an analogous extron (one coding for a protein similar to P) in a slightly lower species.

> > > The biggest improvements might come about by transferring the meta-heuristics (those introns which deal with other introns, rather than extrons).

> > A much simpler kind of prediction is that messing with introns will affect the % viability of mutant offspring. This may be one of the first experiments to perform, due to its general simplicity.

> > More convincing would be the following: cause organisms to mutate, and then to mutate back, and thridly to mutate in the same way AGAIN. We predict that the third mutation will be MUCH faster than the first one.

> > > Yesterday (Thu., Oct. 12) I asked Doug Brutlag about this particular experiment. Schimke (at Stanford) has done it, and gotten just such results. Also, the length of the DNA increases during the initial learning period, decreases during unlearning -- but NOT all the way back to its original shortness, and then increases again. We guess that the extra residual length is the new heuristic introns.

> > When would X have evolved? In particular, when would we expect something as good as Man to appear on the scene?

> > > This is tough to do theoretically. It might be doable empirically, by building a big AI program which simulated evolution (not purely random mutation, like Fogel's), and which started at some place where SOME introns already existed, and which used them to mutate plausibly.

> > > We must also compute when pure chance might have been expected to generate the first crude heuristics.

> > Another prediction is that various kinds of non-random behavior (i.e., mutations occurring in patterns which can be recognized) will be noticed at the base-sequence and even at the gene level.

> > > Brutlag was startled when I asked if this had been observed, since that's precisely the phenomenon he's investigating now.

> > > We must also compute when pure chance might have been expected to generate the first crude heuristics.

> If the paradigm does seem to be verified, what issues should be investigated?

> > The foremost problem, of course, is the intron "code".

> > > We can use hypotheses about unity and simplicity to

guide our investigations, and to buoy our spirits that the answer is not a convoluted one.

> > > We will look at the changes when a heuristic is transferred to various organisms, and induce what it says.

> > Perhaps even prior to tackling the code itself, we must figure out the mechanism whereby the introns are Evalled.

> > > Closely tied with this is, of course, the programming analogues of the form of the introns.

> > > If they are IF/THEN type rules, what is the interpreter? Is the "IF" part partially or totally specified by position? Is the "THEN" part partially or totally a HISTORY of what the last (last few? all past?) modificiations were?

> > > Are there different types? Do some types correspond to data structures, some to plausibility rules which refer to those data structures, and others to interpreters?

> > > Are the numbers right? It would be tragic to find evidence for the above hypotheses, and yet find that the numbers still said man would come out in 100000000000000000 AD. Or the day after bacteria.

> > > > But it would be more tragic to have conceptualized trans-mutation mechanisms, and yet not check to see that we had gone far enough (i.e., as far as Nature has gone by now) -- and not "too" far.

> If the paradigm seems NOT to be verified, what might we do?

> > The failure is probably due to one of two causes,:

> > Most likely, Nature is not as good a programmer as we in AI are today. In that case, let's go back to idea #1: let's try to design heuristics for plausible and implausible mutations, for recordkeeping, for dealing with (synthesizing, modifying, evaluating) other heuristics. They will have to be non-coding sequences, there will have to be an EVALuation mechanism for obeying them at reproduction-time, etc. Then experiments will have to be designed, in which such sequences are built up and inserted into DNA.

> > Less likey, in fact almost incredible, would be if Nature were already a far superior programmer than we. In that case, quite ironically, the next big idea in AI could come from unravelling whatever mecahnism Nature has already developed for efficiently evolving DNA.

> Can we propose a plausible model for how this all might work?

> > Even if it's poorly motivated by empirical evidence, such an "existence proof" is quite convincing -- and quite common in genetics.

> > > Consider Gamow's early scheme for the genetic code.

> > Let us propose a model which is as close to Eurisko as possible

> > > Some sequence of bases function together as a heuristic

> > > Each such heuristic H is delimited by a telltale base sequence h

> > > Each such hHh group has a particular scope, a domain of relevance

> > > > Thus, "use a repressor/anti-repressor mechanism rather than an induction mechansm" might hold true for a patch of DNA which synthesized the organism's most important enzymes.

> > > > In lieu of Lisp-like pointers, we suggest some more analogic way of indicating the scope of hHh.

> > > > As with AM and Eurisko, a natural way of doing this is to place it just before the relevant referent.

> > > > Some base sequences might serve as parentheses to explicitly demarcate the limits of the scope of the heuristic.

> > > > Please note that heuristics can have as their domains sets of

other heuristics!

> > > Each heuristic H consists of a few pieces of information

> > > > A rating (e.g., how often ANY mutation should be tolerated in the section of DNA that comprises the scope of H)

> > > > A (generalized) change that was tried in the past and worked

> > > > > > What the state was before the change

> > > > > > We presume that the state now is the current state

> > > > > > > > At least after the composition of all the H's in sequence

> > > > > > We presume that the change was beneficial

> > > > > > > > Else the new animals would not multiply, and the poor heuristics they possessed would immediately die out (at least, not fix).

> > > > A (generalized) change that was tried in the past and failed

> > > > > > What the state was before the change

> > > > > > We presume that the change was harmful or lethal

> > > > > > > > Else the new animals would have multiplied, and the wrong heuristics that these old animals possess would have slowly died away.

> > > > What is the allowable "language" of actions on the right hand (THEN-) side of each heuristic rule?
One typical action might be gene rearrangement.
WC&W: "It is notable that rates of evolutionary change in gene rearrangement are unusually high in those groups with high rates of phenotypic evolution and speciation."
A related action might be to DUPLICATE a gene; one copy would continue to perform its original function, and the new copy would be available for experimentation.
Other actions might include synthesizing and modifying introns.

> > We should construct a big example scenario of this in action, in detail.

> > > Notation (in addition to the above) must be developed

E = a segment of DNA which translates directly into an enzyme
P = a segment that translates directly into any protein
E(+P) = a segment that translates into an enzyme that increases the rate at which P is produced in the organism/cell.
[...] to denote the scope of heuristics
E(-n%P) + segment translates into enzyme that decreases the production of protein P by about n%.
s = a start or stop sequence (at front or end of P)
More notation about functions of proteins (growth, etc.)

> > > Specify an initial state (for a tiny bit of the nuclein of an organism)

> > > > The sequences that code for various proteins and heuristics
E.g., hH1hhH2h[hH3hhH4hhH5hhH6hhH7h[sP1ssP2s]]
would refer to two protein-encodings, four heuristics relevant to them, and two meta-heuristics relevant to those last four.

> > > > Each Hi and Pi must then be defined in terms of the above notation (e.g., we might say that P1 = E(P3)) or in English.

> > > Go through the simulation

> > > > Look at the various kinds of mutations that might form, and the probabilities of each, and their utilities. Compare with random.

> > > > Include here at least a few cases where heuristics, not merely protein-encodings, get created and get modified.

> > > > Also at this stage, we should make some guesses about the mechansim for applying the heuristics (for obeying them). The need to come up with a simple molecular explanation is at once pressing (to convince skeptics) and deferrable (since many confirming experiments might be done without the precise mechansim being understood).

APPENDIX: THE CONTEXT

Relevant Existing "Knowledge"

Asterisks (*) indicate "facts" that I believed before the idea was formed,
but which (due to subsequent reading/discussion) I now feel are wrong/unknown
Plusses (+) indicate facts I have learned since the idea was formed.

> Mendelism is accepted absolutely.
> > That is, we are completely determined by our genetic makeup.
* > > > In particular, by our genetic materials AT BIRTH
> > > Changing said genetic materials will alter the genetic makeup
-- and hence the "blueprints" of, the design -- of our offspring

> Evolution in the strict Darwinian sense (i.e., solely via a
series of random mutations, with Natural Selection providing
the test for generate&test improvement) is incapable of
accounting for the presence of, e.g., Man on earth today.

> > Certainly, we do not dispute that natural selection operates
> > > E.g., the adaptation (darkening) of city moths' coloration
> > > E.g., in societal artifactual systems (academia, politics,...)
> > Moreover, we concede that simple natural selection could quite
possibly have preserved each "step" toward Man, had each new
improvement come along and co-existed with less evolved bretheren.
> > Certainly, we do not dispute that random mutations occur
> > > The large number of birth defects each year is sad testimony.
> > > The "numbers" make it clear that nothing more than random
genetic mutation is required to account for the phenomenon
whereby bacteria become resistant to some drug.
> > Moreover, random mutations could account for each "step" to Man
> > > A "step" is what Simon would call a "subassembly" -- a stable
design for an organism which is superior to (hence will be
selected for over) the previous design of that organism.
> > We object to the QUANTITATIVE plausibility of the last " > > "
> > > The order of magnitude of such a "pure hillclimbing" toward
* Man can be estimated to be as large as 10↑(10↑6) years !!
> > > > Many of us (e.g., Knuth) see the need for extreme skepticism
of the doctrine that natural selection of superior random mutants
can account for Man evolving in so short a time.
> > > > The mutation rate per gene per generation is around 10↑-7
\+ > > > > Almost all random mutations are deleterious, or at best neutral.
\+ > > > > And there is a good chance that even an advantageous new allele
will be lost (die out before fixation occurs)
due to fluctuations in its frequency in the population as a whole.
\+ > > > The area of quantitative evolution is currently a hot one
in the sense that many articles are coming out:
> > > > > Some recent articles on sequence evolution are trying to
show, e.g., that proteins needn't have evolved too quickly
(that some of Man's proteins are not much different from yeast's)
> > > > Cavalli-Sforza: "The evolution of brain size in man turns out to
be among the most rapid, if not the most rapid, of known
evolutionary processes." (p. 692 of The Genetics of Human Populations)

He then mentions that this enlargement needn't have been gradual, continuous.

> > > In addition, we must bear in mind that natural selection does not tolerate much curvilinear development.

> > > > I.e., a very complex system (like the double-negative repression-repression system for B-galactosidase) would have had to evolve in steps EACH of which was a positive improvement over the last one.

\+ > > > > > Non-Darwinian theories, e.g., about the fixation of large numbers of neutral mutations, are also emerging lately.

> > > > An extreme of this would be to demand that the entire system evolve in one huge simultaneous mutation. Simon shoots this down well in his Science of the Artificial.

\+ > > There are several anomalies in the data about evolution, besides the previous one (the doubt about the RATE of evolution)

> > > Why did man's brain evolve so rapidly?

> > > Why do some proteins evolve at rates 10 times as slow as others?

> > > > Older proteins seem to undergo (on average) a smaller no. of changes

> > > > Some parts of a protein (some amino acids, usually about 5%) are absolutely stable (NEVER appear to have undergone substitution, even during long evolutionary time periods. (Cavalli p.741)

> > > Why is the mutation rate per gene proportional to the total length of the DNA molecule, not a constant? (ABC paper)

\+ > > > Also, there are many riddles presented in articles in Duncan & Weston-Smith's Encyclopedia of Ignorance:

> > > > The Sources of Variation in Evolution (Roy J. Britten)

"How is it possible for future evolutionary flexibility to be preserved when the exigencies of survival apply strong immediate selection pressure? ... Is it simply chance that some species preserve evolutionary flexibility while others do not?... All of these questions suggest that natural selection is a subtle process and that a significant part of the genetic information may not be subject to short-term selection. How could such information be sotred, and over what period of time is it effectively selected? There are aspects of the fossil record which suggest parallel evolution of species lines that have been long separate. Such convergent or parallel evolution does not have an easy explanation and also suggests long-term storage of genetic information. On a molecular level there are also suggestions of freedom from selection pressure, or longer peridos of intgration. For example, mammals contain enough DNA per cell to code for an excessive number of potential genes (though most of this DNA is surely something other than structural genes... There is obviously a lot of DNA in the genome of higher organisms that we can not account for. This has been termed the C-value paradox. To add to the mystery, most of the single copy DNA in primates changes so rapidly in evolution that it is probably under little or no selection pressure. We do not know what unexpressed potentialities exist in all of this 'extra' DNA."

"We have found that a typical gene contains about three-quarters single copy DNA, and about one-quarter sequences present [repeated] in 100 to 10,000 copies int he DNA of a single cell. The individual repeats are more or less imperfect and copies differ by as much as 10 to 20 per cent of their bases."

"1500-15000 significant changes incorporated, after selection, into human DNA in 15 million years. Are these few base substitutions incorporated in the DNA enough to be the source of variation for the last 15 million years of evolution? It seems unlikely unless they had just the right effect. We can think int erms of changes in the gene regulatory system that would affect the form or function of an organ. But how many base substitutions can have such effects? Amino acid substitutions in typical proteins -- no way. Even billions [of small biochemical changes] might not be enough."

> > > > The Edge of Evolution (J.C. Lacey, A.L. Weber, and K.M. Pruitt)
"The primary DNA information, although inside the cell, now represents part of the environment for selecting the super [meta-level] information."
Also: their citation of E. Zuckerland and L. Pauling's "Molecules as documents of evolutionary history", J. Theor. Biol., 8, 357-66, 1965.
> > > > Fallacies of Evolutionary Theory (E.W.F. Tomlin)
"Evolution was an hypothesis which hardened into dogma before it had been thoroughly analysed." "Even sophisticated Darwinians such as Konrad Lorentz assume without question that the origin and formation of species can be explained as a succession of fortuitous variations and mutations passing through the mesh of selection. The oddity of this theory is partially concealed by its mode of presentation." Our tools -- both external ones like rotary saws and internal ones like enzymes -- must have developed "thematically; they cannot have come into being by a series of mutations or mechanical faluts of copying".
> > > > The limitations of evolutionary Theory (John Maynard Smith)

"Suppose that at a time 200 million years ago, during the age of reptiles, some evnt had taken place which doubled the rate of gene mutation in all existing organisms... Would the present state have been reached in only 100 million years? Or would the rate of evolution have stayed much the same?... The short answer is that we do not know. ... A theory of evolution which cannot predict the effect of doubling one of the major parameters of the process leaves something ot be desired."
Enzymes correct the copying errors; since the enzymes are produced by genes, the mutation rate is under genetic control.

> > As an analogue, consider the construction of a large program
> > > Which after all is what DNA is
> > > One might try to randomly change a program, and to (occasionally) randomly add a random new instruction.
> > > It's feasible to synthesize very short programs by such tactics
> > > > PW1 by myself (Green et al. AI Memo 1974)
> > > > Early IBM work on automatic programming (circa 1960)
> > > This method breaks down rapidly as program size/complexity rise
> > > > Small random changes in a complex program (e.g., in assembly language) are usually fatal, almost never beneficial.
> > > > For the obvious combinatorial reasons
> > > > See Fogel et al.'s work on simulated evolution of automata
> > > > > Note his initial success followed by swamping failure
> > > > See also the various Cognitive simulations of neonates
> > > > > John Burge, MIT efforts, etc.
> > > Note that we are not demanding the sui generis synthesis of a large program all in one step
> > > > Like a monkey at a typewriter
> > > > Rather, we are willing to grant as "islands" ANY partial programs which are in ANY I/O way superior to their parents
> > > > > They run faster
> > > > > They use up less space
> > > > > They can do one more tiny thing than their parents
> > > > > (BUT: what about "They produce better mutant offspring [on the average] than their parents do"?)
> > > > > "Any I/O way" means any PHENOTYPE difference.
> > > > Even so, we claim, random mutation is not an effective method from which intelligent programs would evolve.
> > > > > This is the conclusion reached by the above projects which tried such experiments, as well as

the combinatorial conclusion.

> Natural selection is accepted completely
> > Survival of the fittest, in a harsh environment, is the sole criterion for judging improvement
> > > At least in pre-Man ages, which is what we're considering
> > Natural selection is omnipresent and severe
> > > At least, for pre-Man ages.
> > > So, e.g., curvilinear progress is rarely tolerated
> > > > That is, when a mutation produces an inferior animal
> > > > But a mutation generations later combines with the first to result in a distinctly superior species.

> Eurisko is assumed to be viable
> > Not the program, the overall idea
> > This is a somewhat shaky assumption
> > > It is underconditioned by DIRECT empirical verification
> > > > I.e., the program doesn't run yet
> > > But it is plausible in light of AM and other HPP work
> > The idea is the conjunction of the following:
> > > (HPP) Complex tasks call for expert programs
> > > > To construct an expert program, we must somehow put "expertise" into programs.
> > > > Heuristic if-then rules are a reasonable language in which to state (and incorporate) such expertise.
> > > > In particular, Generate&Test alone is much too weak to give adequate performance in complex domains.
> > > (HPP) Heuristic rules can efficiently guide huge searches
> > > (AM) The above applies to exploration which is open-ended research
> > > > At least, in the realm of elementary math theory formation
> > > (EUR) The above applies to "heuristics" as well as "math concepts"
> > > > In fact, a body of heuristics can improve and expand "itself"
> > > > The most simple. elegant, natural, compact, unifying,... way to effect this is merely to represent each heursitic as an object in the domain of the body of heuristics
> > > > > In case the heuristics are like AM's, this means coding each one as a frame-like AM "concept".
> > > > > So, e.g., any heuristic which can generalize the Defin slot of any concept, can generalize the Defin of any heuristic (including, incidentally, itself!)

> DNA is viewable as a program...

> > Transfer RNA "swaps in" the DNA "program", and at the ribosomes it is "EVAL'ed" (messenger RNA brings the required types of "freelist cells"). The "output" is a polypeptide chain (protein).
> > The famous "genetic code" is the key with which triples of base pairs are converted into amino acids. That is the programming language's basic "Print" statement.
> > Simple loop termination (and other regulatory actions) are brought about by the program -- the DNA -- synthesizing certain proteins (which we call enzymes) which are capable of interfering

with the executive control structure (e.g., halting the messenger RNA from reading some parts of the DNA, causing it to start reading from a new place, etc.)

> ... but some subroutines serve as-yet unknown purposes.

> > In higher organisms' DNA, there are many long subsequences which do not appear to be translated (or even translatable) into proteins. They are called "introns", and their biological function is unknown and currently quite a hot topic of speculation.

* > > The percentage of such "non-coding" segments increases as one ascends the evolutionary ladder.

\+ > > > In prokaryotes, there is no trace of extraneous DNA.

\+ > > > In yeast, the simplest eukaryotic organism studied extensively, there is suggestive evidence for a minute amount of introns.

\+ > > > In chick albumen, there is a nontrivial amount of introns.

> > > > This came as quite a shock to researchers, who had previously assumed that all DNA was "extrons" -- that is, codings for proteins.

> > > > The mechanism for ignoring the introns is effected somehow by mRNA, which simply cleaves off introns and leaves extrons as it's copying, before it moves out to a ribosome.

> > > [here, add various experimental results about introns]

\+ > > > Thus there is at present only weakly corraborative evidence for my phylogenetic assumption about the increase in introns.

APPENDIX: A few references

Abrahamson, Seymour, Michael A. Bender, and Alan D. Conger -- ABC -- and Sheldon Wolff, "Uniformity of Radiation-induced Mutation Rates among Different Species", Nature, 245:5246, 460-2, October 26, 1973.

Bukhari, A. I., J.A. Shapiro, and S.L. Adhya, "DNA Insertions, Elements, Plasmids, and Episomes", Cold Spring Harbor Laboratory, 1977.

Cavalli-Sforza, L.L., and W.F. Bodner, "The Genetics of Human Populations", W. H. Freeman and Company, San Francisco, 1971.

Duncan, Ronald, and Miranda Weston-Smith, "The Encyclopedia of Ignorance: Everything you ever wanted to know about the unknown", Pergamon Press, New York, 1977, 205-411.

Wilson, Allan C., Steven S. Carlson, and Thomas J. White, "Biochemical Evolution", Am. Rev. Biochem., 1977, 46:573-639.

Information about introns came through informal discussions with Jerry Feitelson and Doug Brutlag.